Hepatoprotective effects of systemic ER activation

ChIPseq/Epigenome genome - Quantification of reads in diffbound promoters and enhancers

Christian Sommerauer & Carlos Gallardo

08 December, 2023

This script has two parts:

in Part 1, we use thr generated bed file with 182 diffbound promoters (one TSS per gene; the file after diffbind analysis still has multiple TSS per gene, 184 total). This bedfile is then converted into a SAF file using awk in command line. This SAF file, together with BAM files is used for featurecounts to generate a count matrix with all reads in these peak regions (this is analyzed on UPPMAX). The same is done for the 1,816 enhancer peaks.

in Part 2, we import this count matrix into R again, normalize it, calculate significances and generate the plots.

Part 1

```
#The enhancer and promoter (with unique TSS, generated in the chunk above) bed files were used to creat

awk 'OFS="\t" {print $1"."$2"."$3, $1, $2, $3, "."}' in.bed > out.saf

#The saf file was then used to annotate the reads, all bam files (Remove duplicates, blocklist, sorted,
#subread version 2.0.0 was used.

Following featureCounts chunk was used:
featureCounts \
        -g gene_id \
        -s 2 \
        -T 2 \
        -M \
        -F SAF \
        -O \
        -C \
        -a ${SAF_PATH}/example.saf  \
        -o ${OUTPUT_PATH}/example.readCount \
        ${BAM_PATH}/*H3K27ac*MkDup.bam \
        &> ${OUTPUT_PATH}/example.readCount.log

the code was then uploaded to uppmax, features counted, then downloaded, finally the header cleaned up
```

Part 2

Import the reads in peaks which was created using feature counts. Then, normalize the reads to do the analysis. Perform an Anova to compare significances between the conditions.

```
library(tidyverse)
counts_prom <- read.delim("results/Epigenome_analysis/DAc_promoters_142_H3K27ac.clean.readCount", heade
  dplyr::select("CK0744_H3K27ac_CD2", "CD6_H3K27ac_S2_R1_001", "CK0745_H3K27ac_CD9",
         "CK0746_H3K27ac_HFD3", "CK0747_H3K27ac_HFD4", "HFD6_H3K27ac_S8_R1_001",
         "CK0748_H3K27ac_DPN2", "CK0749_H3K27ac_DPN3", "DPN6_H3K27ac_S15_R1_001",
         "DIP3_H3K27ac_S9_R1_001",  "DIP6_H3K27ac_S10_R1_001", "DIP10_H3K27ac_S11_R1_001",
         "E2_2_H3K27ac_S16_R1_001", "CK0750_H3K27ac_E2_8", "CK0751_H3K27ac_E2_9",
         "PPT1_H3K27ac_S12_R1_001",  "PPT2_H3K27ac_S13_R1_001", "PPT3_H3K27ac_S14_R1_001")

names(counts_prom) <- c("CDm2","CDm6", "CDm9","HFDm3","HFDm4","HFDm6", "DPN2","DPN3","DPN6","DIP3","DIP
colsums_prom <- colSums(counts_prom[,])

#normalise per depth
counts_prom_norm <- sweep(counts_prom, 2, colsums_prom, FUN = "/")
counts_prom_norm2 <- counts_prom_norm *10^6
colSums(counts_prom_norm2[,])
```

First, for promoters.

```
##  CDm2  CDm6  CDm9 HFDm3 HFDm4 HFDm6  DPN2  DPN3  DPN6  DIP3  DIP6 DIP10  E2_2
## 1e+06 1e+06 1e+06 1e+06 1e+06 1e+06 1e+06 1e+06 1e+06 1e+06 1e+06 1e+06 1e+06
##  E2_8  E2_9  PPT1  PPT2  PPT3
## 1e+06 1e+06 1e+06 1e+06 1e+06
```

```
#Combine the replicates
counts_prom_norm3 <- counts_prom_norm2 %>%
  mutate(avg_CD = rowMeans(counts_prom_norm2[,1:3])) %>%
  mutate(avg_HFD = rowMeans(counts_prom_norm2[,4:6])) %>%
  mutate(avg_DPN = rowMeans(counts_prom_norm2[,7:9])) %>%
  mutate(avg_DIP = rowMeans(counts_prom_norm2[,10:12])) %>%
  mutate(avg_E2 = rowMeans(counts_prom_norm2[,13:15])) %>%
  mutate(avg_PPT = rowMeans(counts_prom_norm2[,16:18])) %>%
  tibble::rownames_to_column("loc_ID")

prom_enh_diffbound <- readRDS("results/Epigenome_analysis/annotated_diffbind_and_genomewide_promoters_e

Prom_Diffbind_HFDup <- prom_enh_diffbound$promoters_HFDup %>% mutate("loc_ID" = paste0(seqnames, ".",sta
Prom_Diffbind_HFDdown <- prom_enh_diffbound$promoters_HFDdown %>% mutate("loc_ID" = paste0(seqnames, "."

counts_prom_HFDup <- counts_prom_norm3 %>% filter(loc_ID%in%Prom_Diffbind_HFDup$loc_ID) %>% dplyr::selec


counts_prom_HFDup2 <- counts_prom_HFDup %>%
  pivot_longer(cols=2:7) %>% mutate("updown"= "HFDup") %>% group_by(name)

# STATISTICS - One-sided anova.
```

```r
counts_prom_HFDup.stat <- aov(value ~ name, data = counts_prom_HFDup2)
summary(counts_prom_HFDup.stat)
```

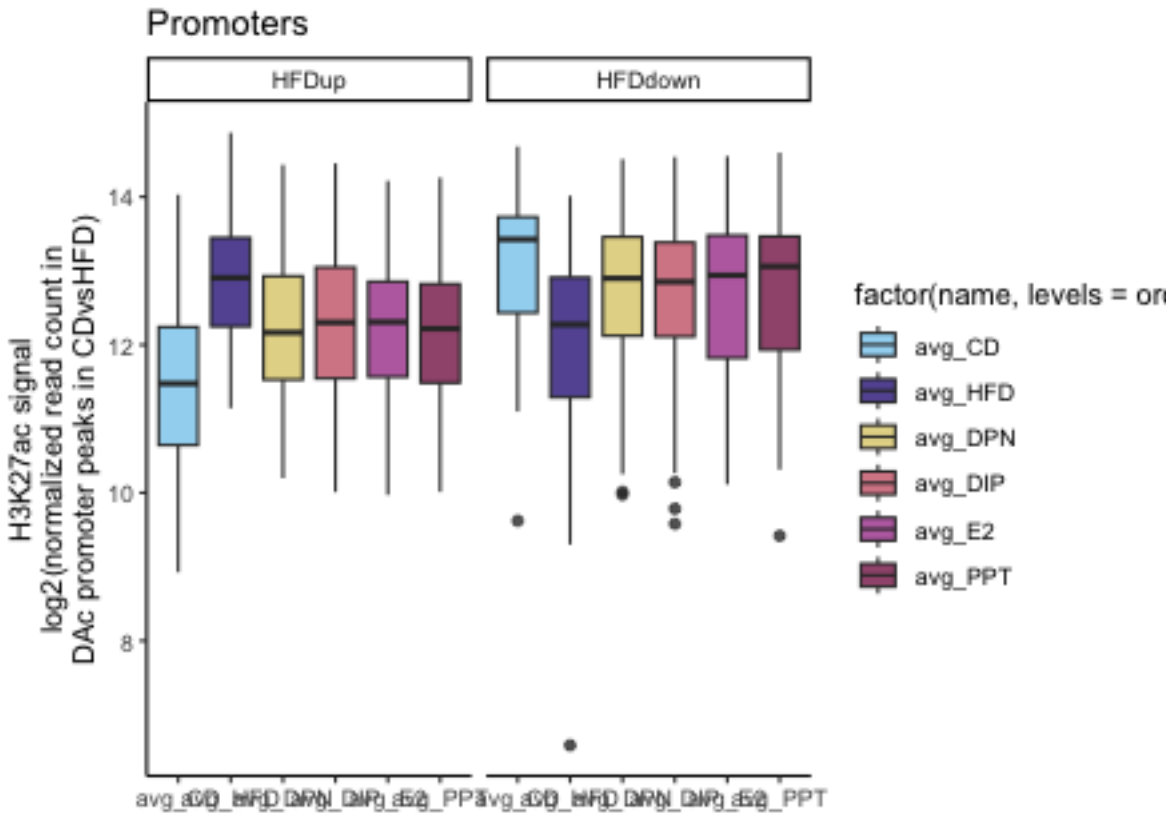
```
##               Df    Sum Sq   Mean Sq F value   Pr(>F)
## name           5 9.863e+08 197255871   11.99 7.45e-11 ***
## Residuals    414 6.812e+09  16453050
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
counts_prom_HFDup.stat1 <- TukeyHSD(counts_prom_HFDup.stat)
counts_prom_HFDup.stat2 <- counts_prom_HFDup.stat1$name

counts_prom_HFDdown <- counts_prom_norm3 %>% filter(loc_ID%in%Prom_Diffbind_HFDdown$loc_ID) %>% dplyr::
counts_prom_HFDdown2 <- counts_prom_HFDdown %>%
  pivot_longer(cols=2:7) %>% mutate("updown"="HFDdown") %>% group_by(name)

# STATISTICS - One-sided anova.
counts_prom_HFDdown.stat <- aov(value ~ name, data = counts_prom_HFDdown2)
summary(counts_prom_HFDdown.stat)
```

```
##               Df    Sum Sq   Mean Sq F value   Pr(>F)
## name           5 9.589e+08 191776541   8.798 5.86e-08 ***
## Residuals    426 9.286e+09  21797457
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
counts_prom_HFDdown.stat1 <- TukeyHSD(counts_prom_HFDdown.stat)
counts_prom_HFDdown.stat2 <- counts_prom_HFDdown.stat1$name

counts_prom_HFD_plot <- rbind(counts_prom_HFDup2, counts_prom_HFDdown2)
counts_prom_HFD_plot$value <- counts_prom_HFD_plot$value+1
```

```r
order <- c("avg_CD", "avg_HFD", "avg_DPN","avg_DIP", "avg_E2", "avg_PPT")
counts_prom_HFD_plot$updown <- factor(counts_prom_HFD_plot$updown, levels = c("HFDup", "HFDdown"))
ggplot(counts_prom_HFD_plot, aes(x=factor(name, levels=order), y=log2(value), fill=factor(name, levels=
  geom_boxplot(alpha=0.8) +
  theme_classic() +
  facet_wrap(vars(updown)) +
  xlab("") +
  ggtitle("Promoters") +
  ylab("H3K27ac signal\n log2(normalized read count in \nDAc promoter peaks in CDvsHFD)") +
  scale_fill_manual(values = c("#88CCEE",  "#332288", "#DDCC77", "#CC6677", "#AA4499",  "#882255"))
```



Plot for promoters

```
library(dplyr)
library(tidyr)
counts_enha <- read.delim("results/Epigenome_analysis/DAc_enhancers_2181_H3K27ac.clean.readCount", head
  dplyr::select("CK0744_H3K27ac_CD2", "CD6_H3K27ac_S2_R1_001", "CK0745_H3K27ac_CD9",
         "CK0746_H3K27ac_HFD3", "CK0747_H3K27ac_HFD4", "HFD6_H3K27ac_S8_R1_001",
         "CK0748_H3K27ac_DPN2", "CK0749_H3K27ac_DPN3", "DPN6_H3K27ac_S15_R1_001",
         "DIP3_H3K27ac_S9_R1_001",  "DIP6_H3K27ac_S10_R1_001", "DIP10_H3K27ac_S11_R1_001",
         "E2_2_H3K27ac_S16_R1_001", "CK0750_H3K27ac_E2_8", "CK0751_H3K27ac_E2_9",
         "PPT1_H3K27ac_S12_R1_001",  "PPT2_H3K27ac_S13_R1_001", "PPT3_H3K27ac_S14_R1_001")

names(counts_enha) <-  c("CDm2","CDm6", "CDm9","HFDm3","HFDm4","HFDm6", "DPN2","DPN3","DPN6","DIP3","DI
colsums_enha <- colSums(counts_enha[,])

counts_enha_norm <- sweep(counts_enha, 2, colsums_enha, FUN = "/")
counts_enha_norm2 <- counts_enha_norm *10^6
colSums(counts_enha_norm2[,])
```

Then, for enhancers

```
##  CDm2  CDm6  CDm9 HFDm3 HFDm4 HFDm6  DPN2  DPN3  DPN6  DIP3  DIP6 DIP10  E2_2
## 1e+06 1e+06 1e+06 1e+06 1e+06 1e+06 1e+06 1e+06 1e+06 1e+06 1e+06 1e+06 1e+06
##  E2_8  E2_9  PPT1  PPT2  PPT3
## 1e+06 1e+06 1e+06 1e+06 1e+06
```

```r
#Combine the replicates
counts_enha_norm3 <- counts_enha_norm2 %>%
  mutate(avg_CD = rowMeans(counts_enha_norm2[,1:3])) %>%
  mutate(avg_HFD = rowMeans(counts_enha_norm2[,4:6])) %>%
  mutate(avg_DPN = rowMeans(counts_enha_norm2[,7:9])) %>%
  mutate(avg_DIP = rowMeans(counts_enha_norm2[,10:12])) %>%
  mutate(avg_E2 = rowMeans(counts_enha_norm2[,13:15])) %>%
  mutate(avg_PPT = rowMeans(counts_enha_norm2[,15:18])) %>%
  tibble::rownames_to_column("loc_ID")

Enha_Diffbind_HFDup <- prom_enh_diffbound$enhancers_HFDup %>% mutate("loc_ID" = paste0(seqnames, ".",sta
Enha_Diffbind_HFDdown <- prom_enh_diffbound$enhancers_HFDdown %>% mutate("loc_ID" = paste0(seqnames, "."

counts_enha_HFDup <- counts_enha_norm3 %>% filter(loc_ID%in%Enha_Diffbind_HFDup$loc_ID) %>% dplyr::sele
counts_enha_HFDup2 <- counts_enha_HFDup %>%
  pivot_longer(cols=2:7) %>% mutate("updown"= "HFDup")

# STATISTICS - One-sided anova.
res.aov.enha.HFDup.stat <- aov(value ~ name, data = counts_enha_HFDup2)
summary(res.aov.enha.HFDup.stat)
```

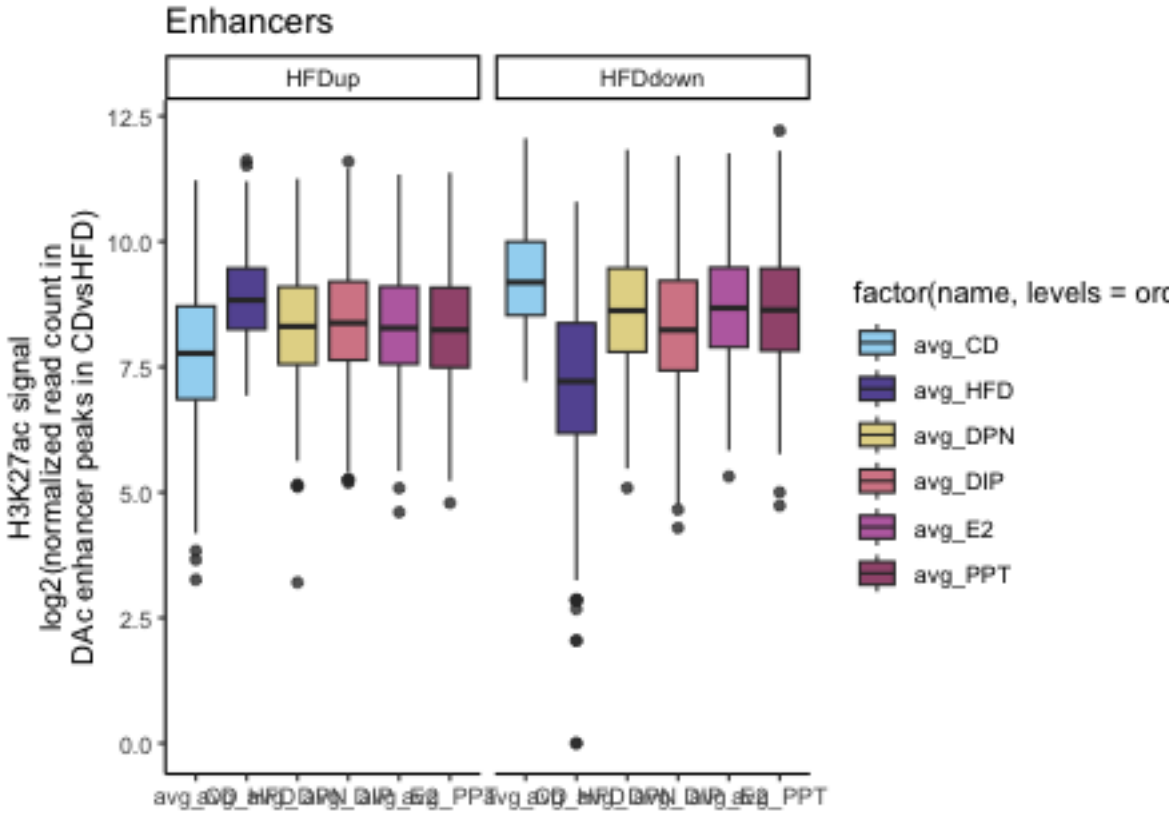
```
##               Df    Sum Sq Mean Sq F value Pr(>F)
## name           5  47811754 9562351   86.76 <2e-16 ***
## Residuals   8994 991236041  110211
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
res.aov.enha.HFDup.stat1 <- TukeyHSD(res.aov.enha.HFDup.stat)
res.aov.enha.HFDup.stat2 <- res.aov.enha.HFDup.stat1$name

counts_enha_HFDdown <- counts_enha_norm3 %>% filter(loc_ID%in%Enha_Diffbind_HFDdown$loc_ID) %>% dplyr::s
counts_enha_HFDdown2 <- counts_enha_HFDdown %>%
  pivot_longer(cols=2:7) %>% mutate("updown"="HFDdown")

# STATISTICS - One-sided anova.
res.aov.enha.HFDdown.stat <- aov(value ~ name, data = counts_enha_HFDdown2)
summary(res.aov.enha.HFDdown.stat)
```

```
##               Df    Sum Sq  Mean Sq F value Pr(>F)
## name           5 105312233 21062447   86.72 <2e-16 ***
## Residuals   4080 990931503   242875
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
res.aov.enha.HFDdown.stat1 <- TukeyHSD(res.aov.enha.HFDdown.stat)
res.aov.enha.HFDdown.stat2 <- res.aov.enha.HFDdown.stat1$name

counts_enha_HFD_plot <- rbind(counts_enha_HFDup2, counts_enha_HFDdown2)
counts_enha_HFD_plot$value <- counts_enha_HFD_plot$value+1
```

```
order <- c("avg_CD", "avg_HFD", "avg_DPN","avg_DIP", "avg_E2", "avg_PPT")
counts_enha_HFD_plot$updown <- factor(counts_enha_HFD_plot$updown, levels = c("HFDup", "HFDdown"))
ggplot(counts_enha_HFD_plot, aes(x=factor(name, levels=order), y=log2(value), fill=factor(name, levels=
  geom_boxplot(alpha=0.8) +
  theme_classic() +
  facet_wrap(vars(updown)) +
  xlab("") +
  ggtitle("Enhancers") +
  ylab("H3K27ac signal\n log2(normalized read count in \nDAc enhancer peaks in CDvsHFD)") +
  scale_fill_manual(values = c("#88CCEE",  "#332288", "#DDCC77", "#CC6677", "#AA4499",  "#882255"))
```



Plot for enhancers

```
sessionInfo()
```

```
## R version 4.2.3 (2023-03-15)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur ... 10.16
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.2/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.2/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
```

```
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] lubridate_1.9.2 forcats_1.0.0   stringr_1.5.0   dplyr_1.1.2
##  [5] purrr_1.0.2     readr_2.1.4     tidyr_1.3.0     tibble_3.2.1
##  [9] ggplot2_3.4.3   tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
##  [1] highr_0.10        pillar_1.9.0      compiler_4.2.3    tools_4.2.3
##  [5] digest_0.6.33     timechange_0.2.0  evaluate_0.21     lifecycle_1.0.3
##  [9] gtable_0.3.3      pkgconfig_2.0.3   rlang_1.1.1       cli_3.6.1
## [13] rstudioapi_0.15.0 yaml_2.3.7        xfun_0.39         fastmap_1.1.1
## [17] withr_2.5.0       knitr_1.43        generics_0.1.3    vctrs_0.6.3
## [21] hms_1.1.3         grid_4.2.3        tidyselect_1.2.0  glue_1.6.2
## [25] R6_2.5.1          fansi_1.0.4       rmarkdown_2.23    farver_2.1.1
## [29] tzdb_0.4.0        magrittr_2.0.3    scales_1.2.1      htmltools_0.5.5
## [33] colorspace_2.1-0  labeling_0.4.2    utf8_1.2.3        stringi_1.7.12
## [37] munsell_0.5.0
```